AI BASED DISCOURSE FOR BANKING INDUSTRY

ABSTRACT

Artificial Intelligence enables banks to manage huge volumes of data at record speed to derive valuable insights from it. Features such as AI bots, digital payment advisers and biometric fraud detection mechanisms lead to higher quality of services to a wider customer base. Knowing the customer churn rate is a key indicator for any business. According to a study by Bain & Company, improving the customer retention rate for existing customers by just 5 percent can improve a company's profitability by 25 to 95 percent.

The number of service providers are being increased very rapidly in every business. In these days, there is no shortage of options for customers in the banking sector when choosing where to put their money. As a result, customer churn and engagement has become one of the top issues for most of the banks. In this project, a method to predicts the customer churn in a Bank, using machine learning techniques, which is a branch of artificial intelligence is proposed. This project promotes the exploration of the likelihood of churn by analyzing customer behavior. The KNN, SVM, Decision Tree, and Random Forest classifiers are used in this project. Also, some feature selection methods have been done to find the more relevant features and to verify system performance. The experimentation was conducted on the churn modeling dataset from Kaggle. The results are compared to find an appropriate model with higher precision and predictability. As a result, the use of the Random Forest model after oversampling is better compared to other models in terms of accuracy.

It is very easy for customers to switch from one organization(Bank) to another for a better service quality or price rates. Organizations are convinced that recruiting new customers is far more expensive and hard than keeping existing clients. But delivering reliable service on time and in budget to customers while maintaining a good working partnership with them is another significant challenge for them. They need to consider consumers and their needs to resolve these challenges. Among these, one of their primary emphasis will be on client churn. Customer churn takes place when clients or subscribers cease to engage incorporation with a company or service. For any organization, winning business from new clients means going via the sales pipeline, using their sales and marketing assets in the cycle. Customer retention, on the other hand, is usually more budget-effective, because they have already gained the confidence and loyalty of current customers. So, the need for a system that can efficiently predict customer churn in the early stages is really important for any organization. This paper aims to build a framework

that can predict the client churn in the banking sector using some Machine learning techniques.

LITERATURE SURVEY

- M.A.H. Farquad proposed a hybrid approach to overcome the drawbacks of general SVM model which generates a black box model (i.e., it does not reveal the knowledge gained during training in human understandable form). The hybrid approach contains three phases: In the first phase, SVM-RFE (SVM-recursive feature elimination) is employed to reduce the feature set. In the second phase, dataset with reduced features is then used to obtain SVM model and support vectors are extracted. In the final phase, rules are then generated using Naive Bayes Tree (NBTree which is combination of Decision tree with naive Bayesian Classifier). The dataset used here is bank credit card customer dataset (Business Intelligence Cup 2004) which is highly unbalanced with 93.24% loyal and 6.76% churned customers. The experimental showed that the model does not scalable to large datsets.

- Chih-Fong Tsai introduced the hybrid neural networks techniques to predict the customer churners in a CRM dataset provided by American telecom companies. Here, they built two hybrid models by combining two different neural network techniques like back-propagation artificial neural networks (ANN) and self organizing maps (SOM) for churn prediction. The hybrid models are ANN combined with ANN (ANN + ANN) and SOM combined with ANN (SOM + ANN). In particular, the first method of the two hybrid models performs the data reduction task by filtering out unrepresentative training data.

- Wouter Verbeke proposed the application of Ant-Miner+ and ALBA algorithms on a publicly available churn prediction dataset in order to build accurate as well as comprehensible classification rule-sets churn prediction models. Ant-Miner+ is a high performing data mining method based on the principles of Ant Colony Optimization which allows to include domain knowledge by imposing monotonicity constraints on the final rule-set. The advantages of Ant-Miner+ are high accuracy, comprehensibility of the generated models and the possibility to demand intuitive predictive models. Active Learning Based Approach (ALBA) for SVM rule extraction is a rule extraction algorithm, which combines the high predictive accuracy of a non-

linear support vector machine model with the comprehensibility of the ruleset format.

- Ning Lu proposed the use of boosting algorithms to enhance a customer churn prediction model in which customers are separated into two clusters based on the weight assigned by the boosting algorithm. As a result, a high risky customer cluster has been found. Logistic regression is used as a basis learner, and a churn prediction model is built on each cluster, respectively. The experimental results showed that boosting algorithm provides a good separation of churn data when compared with a single logistic regression model.

- Benlan He suggested a customer churn prediction methodology based on SVM model, and used random sampling method to improve SVM model by considering the imbalance characteristics of customer data sets. A support vector machine constructs a hyper-plane in a high- or infinitedimensional space, which can be used for classification. Random sampling method can be used to change the distribution of data in order to reduce the imbalance of the dataset. Imbalance in dataset is caused due to the low proportion of churners.